



McGill

Department of Epidemiology, Biostatistics & Occupational Health
McGill University
1020 Pine Avenue West
Montreal, Quebec
H3A 1A2
Canada

July 24, 2009

Scientific and Archive Staff
Berry Bros. & Rudd

Dear Sir/Madam

I am writing to you, first to ask about some historical scientific documents that were generated at your company in the early 1880s, and second to explore the possibility of a 21st century scientific study involving the same data-source.

Let me introduce myself, and the circumstances leading up to these requests. I am a statistician and professor in the faculty of Medicine at McGill University. I began my career as a statistician involved with clinical trials of cancer therapy in the U.S., and since joining McGill in 1980, I have collaborated on a wide range of epidemiologic research (there are links to my resume and publications on the left hand side of my homepage <http://www.epi.mcgill.ca/hanley>). Recently, I have carried out some archival research, partly for scientific reasons, and partly as a way to stimulate interest in statistical methods, and the study of the history of statistical methods and of public health.

In 2001 we uncovered in the archives at UCL some data on heights of successive generations of families that Francis Galton, a cousin of Charles Darwin, had used when he discovered and documented his famous “regression to mediocrity” law that underlies much of natural inheritance (these data are now accessible on my webpage, and the article describing them is available under the reprints tab). Galton (see <http://www.galton.org/>) was a remarkable 19th century gentleman scholar who turned his hands and mind to many topics, such as statistical methods, geography, meteorology [he discovered anti-cyclones and devised the first weather maps], psychology, anthropometry, fingerprinting, and genetics. While some will consider that his interest in genetics was misguided, no one denies his prodigious scientific output in all of these diverse areas.

One of his investigations – on a topic even more relevant today – addressed how successive generations “put on weight” and what the life course pattern of weight gain/loss is. And it involved what he called a “quarry” of data that your company had assembled over the previous century. The scientific paper he produced is not well known. Since I suspect that even your own company is not aware of it, I am attaching a copy, along with the press attention it got on this side of the Atlantic at the time, and the two pages from Karl Pearson’s 1930 biography of Galton, where he tells us that even he, a good friend of Galton’s in the 1890s and 1900s, was unaware of it.

This brings me to my *first* request: in the second column of p. 266, just above Fig. 2, Galton tells us that his statistical analyses were based on 139 schedules that Messrs. Berry had provided him. Since Galton was an obsessive keeper of raw material, I had fully expected to find the originals among the 200 boxes of archived materials at UCL, but I do not find any mention of them in the detailed catalogue of the archive. I had inspected some of the archived material myself in 2005 for other purposes and did not come across it, but I will look more diligently when I again visit the archives at the end of next month. However, from what Pearson says in his appendix to Ch. XIV, I think my chances of finding these 139 completed schedules at UCL are small.

That these material seem lost is a pity, since there is now considerable interest in “life-course epidemiology (e.g., Sir Michael Marmot at UCL has been following British civil servants in the Whitehall study of health and social well being) and in how to statistically analyze such longitudinal data. And just as with the data Galton collected on stature, it is always good to motivate our epidemiology and statistics students with real, original and interesting data, and to test the newer statistical methods being developed for life course data (for all we know, Galton’s simple methods may turn out to be just as good as the newer fancier ones).

So, I wonder if by any chance Galton used duplicate copies of the schedules, and had Messrs. Berry keep one complete set? He used duplicates when people were measured at the International Health Exhibitionⁱ held in London the following year – each client got the front copy, and I have seen the “carbon” copies at UCL. If your company does have duplicates of the anonymous schedules, their historical nature would considerably enrich the learning experience of students.

I approach my *second* request very delicately, since it probably evokes the same concerns of confidentiality that attended Galton’s work. However, as he did, I expect that there are ways to respect the confidentiality of the data your company had gathered, and yet to use this “quarry” of data for modern public health research. There are few – if any – carefully collected data with the time-span (more than 200 years) or life span (ages 20-70 or more) of this type anywhere else in the world. Yes there are plenty of data on the heights and weights of army recruits over the centuries, but only when they were aged 18. And yes there are a number of cohorts such as Whitehall and Framingham studies, but they only span the last half of the 20th century at most. Your data would allow us to get an unbroken picture over 200-300 years. Other (shorter) nontraditional datasets have been used to study climate change, such as a lottery based on when the ice breaks up in Alaskaⁱⁱ and detailed gardening records of the flowering time of British plantsⁱⁱⁱ.

So, I would like to explore with you the possibility of using some of your ledgers for scientific purposes. What could be done more easily depends on which way the ledgers are kept. If, as I suspect, each person is allotted one line in the ledger, it would be easier to track changes within the same individual (as Galton did). On the other hand, if the

ⁱ His brochure listed where else in London one might be measured, and he mentions Messrs. Berry.

ⁱⁱ Climate Change in Nontraditional Data Sets Raphael Sagarin and Fiorenza Micheli www.sciencemag.org SCIENCE VOL 294 26 OCTOBER 2001 811

ⁱⁱⁱ Rapid Changes in Flowering Time in British Plants A. H. Fitter and R. S. R. Fitter. www.sciencemag.org SCIENCE VOL 296 31 MAY 2002

entries are simply serial in time, it would be easier to track changes in weights over calendar time; it would also remove many of the confidentiality concerns, since we could simply take an anonymous statistical snapshot (a few pages) every 5 or 10 years.

I presume you would have no objection to telling me how the entries are arranged. From there, we could discuss possible uses, procedures for protecting confidentiality, and data extraction^{iv} – the use of digital photographs at the UCL archives made it very efficient, and would avoid having to “beg Messrs. Berry to find a clerk for me” and spend a lot of time on this.

As I wonder why this item of data has had such appeal for your customers, I can think of two reasons. We have little control over our height, and it doesn’t change much over time. Today, given a bathroom scales and a bit of privacy, everyone I know would step on it and weigh himself or herself. Where I grew up in rural Ireland in the 1950s, nobody had such scales; we did have a spring scales for weighing grain etc. When children went into town with their parents a few time a year, the merchants would always measure them on the beam scale and tell their parents how well they were growing.

So, I congratulate Messrs Berry for assembling this long data-series and look forward to discussing with you ways in which this series can again be quarried scientifically. It’s a pity nobody had taken up Galton’s call before, but I hope we can now do so.

I will be in London on Friday August the 28th, and would like to take an appropriately guarded look at the ledgers first hand, and to contribute an entry of my own. In the meantime, I would like to hear from you as to how open Messrs Berry are to some arrangement with me to exploit this resource for an anonymous but useful (in the aggregate) statistical examination. As I am starting out for Europe next week, the best way to reach me between next week and the last week of August is via e-mail.

Yours sincerely,

James A Hanley, PhD

A handwritten signature in black ink that reads "James A Hanley". The signature is written in a cursive, slightly slanted style.

Professor

E-mail: james.hanley@mcgill.ca

^{iv} I could imagine that it would be easier nowadays to have the entries computerized, but I can also imagine that that would take away some of the allure of the historical continuity of the series.